

TEMA 5 CLUSTERING

(Algoritmo BASIC ITERATIVE MINIMUM SQUARED ERROR CLUSTERING)

Ejercicio 1

En el análisis de la media que representa un cluster formado por un conjunto de vectores, demuestre las siguientes variaciones de la media, al mover una muestra de un cluster a otro.

$$\hat{\mathbf{x}} \in D_i \Rightarrow \hat{\mathbf{x}} \in D_j$$

$$\mathbf{m}_j^* = \mathbf{m}_j + \frac{\hat{\mathbf{x}} - \mathbf{m}_j}{n_j + 1}; \quad \mathbf{m}_i^* = \mathbf{m}_i - \frac{\hat{\mathbf{x}} - \mathbf{m}_i}{n_i - 1}$$

$$\text{con } \mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x}_k \in D_i} \mathbf{x}_k; \quad \mathbf{m}_j = \frac{1}{n_j} \sum_{\mathbf{x}_k \in D_j} \mathbf{x}_k$$

Ejercicio 2

En el análisis de la función objetivo al realizar clustering, demuestre que el incremento en la función objetivo $J_j^* - J_j$ que incrementa en la muestra $\hat{\mathbf{x}}$ es:

$$J_j^* = \sum_{\mathbf{x} \in D_j} \|\mathbf{x} - \mathbf{m}_j^*\|^2 + \|\hat{\mathbf{x}} - \mathbf{m}_j^*\|^2 =$$

$$\left(\sum_{\mathbf{x} \in D_j} \left\| \mathbf{x} - \mathbf{m}_j - \frac{\hat{\mathbf{x}} - \mathbf{m}_j}{n_j + 1} \right\|^2 \right) + \left\| \frac{n_j}{n_j + 1} (\hat{\mathbf{x}} - \mathbf{m}_j) \right\|^2 =$$

$$J_j + \frac{n_j}{n_j + 1} \|\hat{\mathbf{x}} - \mathbf{m}_j\|^2$$

Ejercicio 3

En el análisis de la función objetivo al realizar clustering, demuestre que el decremento en la función objetivo $J_j^* - J_j$ que decrementa en la muestra $\hat{\mathbf{x}}$ es:

$$J_i^* = \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i^*\|^2 - \|\hat{\mathbf{x}} - \mathbf{m}_i^*\|^2 =$$

$$\left(\sum_{\mathbf{x} \in D_i} \left\| \mathbf{x} - \mathbf{m}_i + \frac{\hat{\mathbf{x}} - \mathbf{m}_i}{n_i - 1} \right\|^2 \right) - \left\| \frac{n_i}{n_i - 1} (\hat{\mathbf{x}} - \mathbf{m}_i) \right\|^2 =$$

$$J_i - \frac{n_i}{n_i - 1} \|\hat{\mathbf{x}} - \mathbf{m}_i\|^2$$